

RESEARCH ARTICLE

Comparative transcriptomics revealed enhanced light responses, energy transport and storage in domestication of cassava (*Manihot esculenta*)

Zhiqiang XIA^{1,2*}, Xin CHEN^{1,2*}, Cheng LU^{1,2*}, Meiling ZOU^{1,2}, Shujuan WANG^{1,2}, Yang ZHANG^{1,2},
Kun PAN^{1,2}, Xincheng ZHOU^{1,2}, Haiyan WANG^{1,2}, Wenquan WANG (✉)^{1,2}

¹ The Institute of Tropical Bioscience and Biotechnology (ITBB), Chinese Academy of Tropical Agricultural Sciences (CATAS),
Haikou 571101, China

² Key Laboratory of Biology and Genetic Resources for Tropical Crops, Ministry of Agriculture, Haikou 571101, China

Abstract Cassava is a staple food, feed and bioenergy crop important to the world especially in the tropics. Domesticated cassava is characterized by powerful carbohydrate accumulation but its wild progenitor is not. Here, we investigated the transcriptional differences of eight cDNA libraries derived from developing leaf, stem and storage root of cassava cv. Arg7 and an ancestor line, W14, using next generation sequencing system. A total of 41302 assembled transcripts were obtained and from these, 25961 transcripts with FPKM \geq 3 in at least one library were named the expressed genes. A total of 2117, 1963 and 3584 transcripts were found to be differentially expressed in leaf, stem and storage root (150 d after planting), respectively, between Arg7 and W14 and ascribed to 103, 93 and 119 important pathways in leaf, stem and storage root, respectively. The highlight of this work is that the genes involved in light response, such as those for photosystem I (PSA) and photosystem II (PSB), other genes involved in light harvesting, and some of the genes in the Calvin cycle of carbon fixation were specially upregulated in leaf. Genes for transport and also for key rate-limiting enzymes (PFK, PGK and PK, GAPDH) coupling ATP consumption in glycolysis pathway were predominantly expressed in stem, and genes for sucrose degradation (INVs), amylose synthesis (GBSS) and hydrolysis (RCP1, AMYs), the three key steps of starch metabolism, and transport associated with energy translocation (ABC, AVPs and ATPase) and their upstream transcription factors had enhanced expression in storage root in domesticated cassava. Co-expression networks among the pathways in each organs revealed the

relationship of the genes involved, and uncovered some of the important hub genes and transcription factors targeting genes for photosynthesis, transportation and starch biosynthesis.

Keywords cassava, comparative transcriptomics, energy transport, photosynthesis, starch synthesis

1 Introduction

Cassava (*Manihot esculenta*) is an extremely important tropical crop ranked in the top three tuber/root crops along with potato and sweet potato. Its tropical origination makes cassava a unique staple food in tropical and subtropical areas. Over 18 Mhm² of cassava is grown widely distributed from 30° N to S and feeds over 600 million people. Given the food shortage and demand for biofuels, cassava can make an increasingly important contribution as a highly efficient starch crop, because of its high potential photosynthesis and starch accumulation, and strong adaption to drought and barren soils. It is reported that cultivated cassava has a higher rate of photosynthesis than C₃ plants^[1], but its wild progenitor, *M. esculenta* ssp. *flabellifolia*, and other 20 wild relatives showed C₃ rather than C₄ characteristics^[2–7]. The wild ancestor of cassava is a woody shrub that originated in primeval tropical forests. It entered the savannah over the course of its evolution and was domesticated into cultivated cassava as a perennial shrub that can complete its life cycle in one year and can reproduce both sexually and asexually. Its genome is not large but heterozygous, and the draft genomes of cassava and a wild ancestor have been completed^[8]. This will be an essential resource to understand the evolution of the high efficiency of starch accumulation of cassava.

At the level of expressed genes, a cassava cDNA library

Received November 4, 2016; accepted December 14, 2016

Correspondence: wangwenquan@itbb.org.cn

*These authors contribute equally to the work

of bulking root has been reported. This library annotated 2878 unigenes and identified many involved in starch metabolism^[9]. In recent years, cDNA amplified fragment length polymorphisms (cDNA-AFLP) and microarrays have been used to screen differentially expressed transcripts in cassava, such as those for resistance to cassava bacterial blight (CBB) and storage root formation^[10–12] and microarrays have been used for study of resistance to CBB, drought/low-temperature^[13–15] and post-harvest physiological deterioration^[16]. Today, next-generation high-throughput RNA sequencing technique (RNA-Seq) is broadly used for mining global transcription profiles. The transcriptome in rice endosperm at 3, 6 and 10 d after pollination, revealed several genes involved in the accumulation of substances such as starch/sucrose, whereas protein metabolism was upregulated between 3 and 10 d after pollination^[17]. Universal RNA-Seq found *TaIPS1* and 12 candidate genes respond to P_i starvation in wheat^[18] and white lupin^[19], respectively. In cassava, RNA-Seq was used for analysis of mosaic virus-infected susceptible and tolerant landraces and highlighted differences in resistance, basal defense and cell wall associated genes during infection^[20]. Also, transcriptome profiling of low-temperature treated cassava apical shoots showed dynamic responses to cold stress^[21]. Several studies reported the high level expression of genes for photosynthesis, sucrose transport and starch metabolism^[22–24], but there was no analysis of the change during evolution. In this study, we focused on comparison of transcriptional profiling between the wild progenitor *M. esculenta* ssp. *flabellifolia* W14 and cultivated *M. esculenta* cv. Arg7. RNA-Seq was used to investigate the global transcriptome in the leaf, stem and storage root. Differentially expressed transcripts and expression patterns were analyzed and for first time revealed some new concepts for the higher photosynthesis and starch accumulation rates in cultivated cassava compared to its wild progenitor.

2 Materials and methods

2.1 Plant materials

The wild progenitor of cultivated cassava, *M. esculenta* ssp. *flabellifolia* W14 donated by CIAT, was originally collected from the south rim of Amazon basin. One accession of W14 is characterized by lower storage root yield and low starch content in the storage roots (about 5%); *M. esculenta* cv. Arg7 came from Argentina, and is a cultivated variety with high yield and higher starch content in the storage roots (about 30%). These two lines were planted in an experimental station in tropical Hainan Island in 2011. During the growth of plants, 8 samples including leaves and immature stems at 150 DAP (days after planting) of W14 and Arg7, storage roots of W14 at 150

DAP and three samples of storage roots of Arg7 at 90, 150 and 210 DAP were obtained for RNA-Seq analysis. At each time point, samples were taken from three different individuals at 10 am and snap frozen in liquid nitrogen before being stored at –80°C.

2.2 RNA extraction and sequencing

Total RNA was extracted from leaves, immature stems and storage roots with the RNAlant Reagent (Tiangen, Beijing, China) and purified with the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The cDNA libraries were prepared for sequencing with the Illumina HiSeq2000 Sequencer following the protocol of Zhong et al.^[25] and 9.69 to 68.67 million clean reads of approximately 52 bp each were generated for all samples.

2.3 Mapping reads to the reference genome

The cassava genome and gene information were downloaded from Phytozome (<http://www.phytozome.net/cassava.php>). The sequence map files generated by TopHat were used as input for the Cufflinks software^[26], which assembles the alignments in the sequence map file into transcripts. Cufflinks performs this assembly independently from existing gene annotations and constructs a minimum set of transcripts that best describes the RNA-Seq reads, estimating transcript abundance as specific fragments detected per kb of transcript per million mapped reads (FPKM). Cuffmerge was used to merge these eight assemblies with the reference annotation into a single file, which was later used to identify the differentially expressed genes.

2.4 Analysis of differential gene expression

The logarithm of the FPKM is normally distributed and homoscedastic in nature (homogeneity of variance). The R package, DEGseq (<http://bioconductor.org/packages/release/bioc/html/DEGseq.html>) was used to identify differentially expressed genes with a random sampling model based on the read count for each gene by performing Fisher's exact test with robust false discovery rate correction to obtain an adjusted *P*-value between certain test gene groups and the whole genome annotation.

2.5 Annotation of assembled transcripts

The predicted functions of the genes were determined with InterProScan^[27] by comparing with protein databases, such as Pfam, PRINTS, ProFile, SuperFamily, ProDom and SMARTGOIDS for each gene were obtained from the corresponding InterProScan entry. Thereafter, the WEGO software^[28] was used to perform the GO functional classification, with a BLASTP (<ftp://ftp.ncbi.nih.gov/blast/executables/blast±/LATEST/>) parameter of 1×10^{-5}

to search for the best hit for the peptide sequence against genes in the KEGG^[29] database. Within the significant KEGG category, the enrichment Re was given by the formula $Re = (n_f/n)/(N_f/N)$, where n_f is the number of differentially expressed genes within the particular category, n is the total number of genes within the same category, N_f is the number of differentially expressed genes in the entire RNA-Seq data set and N is the total number of genes in the RNA-Seq data set. KEGG pathway network analyses were performed with Cytoscape software (version 2.6.2, <http://www.cytoscape.org>).

Putative functions for the unique sequences were assigned by performing BLASTX queries against the *Arabidopsis* protein database (TAIR10_peptide; <http://www.arabidopsis.org>). The MapMan tool facilitated the classification of transcripts into hierarchical categories in a manner that alleviated the redundancy present in other commonly used ontologies and was used to view the annotated process of interest or metabolic pathway for groups of cassava transcripts.

The R package MINET^[30] was used to construct the gene association network. MINET was developed from the small-sample inference framework of graphical Cytoscape. The mutual information between all pairs of variables in the data set is computed according to the estimator argument. Then the algorithm considers the estimated mutual information in order to build the network.

2.6 Validation of gene expression by quantitative RT-PCR

Four samples, AL and WL, AR150 and WR150, were used to confirm the RNA-Seq data. First, the samples of total RNA were converted to single-stranded cDNA using oligo-dT and reverse transcriptase. Then, a gene-specific primer pair was designed for each gene, after which PCR amplification was carried out using a SYBR Premix Ex TaqTM kit (Takara, Dalian, China) and a Rotor-Gene 6000 Real-Time PCR Machine (Corbett Robotics, Brisbane, Australia). The β -actin gene was used as a control and all PCR amplifications were repeated three times. The relative

expression of the genes between Arg7 and W14 was quantified using the $\Delta\Delta C_T$ method. The primer pairs are listed in Appendix A.

3 Results

3.1 RNA-Seq analysis of developing leaf, stem and storage root of cassava

Under field grown conditions, at three key development stages, leaves, stems and storage roots of Arg7 and W14 were collected and used for RNA-Seq analysis. Eight annotated transcripts were named as AL (Arg7 leaf), AS (Arg7 stem), AR90 (Arg7 storage root at 90 DAP), AR150 (Arg7 storage root at 150 DAP), AR210 (Arg7 storage root at 210 DAP), WL (W14 leaf), WS (W14 stem) and WR150 (W14 storage root at 150 DAP). A total of 126.65 million sequence reads were aligned to the draft cassava genome by TopHat and Cufflinks^[26] (eight libraries have been deposited in GenBank/SRR (Table 1). There were a total of 41302 transcripts assembled, with an average of about 86% of the reads matched to the draft genome. Among them, 25961 were considered to be transcriptionally active with FPKM ≥ 3 in at least one library. Between Arg7 and W14, the number of transcripts ranged from 39863 vs 37822 in all libraries (Appendix B, Fig. S1).

3.2 Comparative gene enrichment analysis between cultivated Arg7 and wild W14

All the transcripts were classified into three main GO categories. Among the assembled transcripts of Arg7 and W14, 10772 and 10735 transcripts, respectively, were assigned to GO categories and they were involved in 360 pathways (Appendix C). Among the GO-categorized transcripts, the main cellular component categories were cell, cell part, organelle, organelle part and macromolecular complex. The primary molecular function categories were binding activity, catalytic activity, transcription

Table 1 Summary transcripts from eight developing organs of Arg7 (*Manihot esculenta* ssp. *flabellifolia*) and W14 (*M. esculenta* ssp. *esculenta*)

Item	Total reads	Mapped reads	Percentage of mapped reads	Number of transcripts	Mean length of transcript/bp	Longest transcript length/bp	N50/bp	GC content
W14 (<i>M. esculenta</i> SSP. <i>flabellifolia</i>)	WL	68673876	51343307	74.76%	41611 37822 41302	2132	59184	4037
	WS	13359172	10628463	79.56%	36778			
	WR150	9693871	7798931	80.45%	33606			
	AL	30710363	27833673	90.63%	35946			
Arg7 (<i>M. esculenta</i> SSP. <i>esculenta</i>)	AS	29605379	26962390	91.07%	36910	39863	2017	73263
	AR90	12034644	10586680	87.97%	33893			
	AR150	15087006	13957544	92.51%	34937			
	AR210	39221907	35604958	90.78%	37050			

regulator activity, transporter activity and structural molecule activity. The main biological process categories were cell process, metabolic process, biological regulation, response to stimulus and pigmentation. In general, for the 47 main pathways, the number of annotated transcripts did not differ markedly between Arg7 and W14. However, there were more annotated transcripts in two pathways (nutrient reservoir activity and locomotion process) in Arg7 than in W14 (Fig. 1).

3.3 Global analysis of gene differential expression profiling

The transcript profiles of the cultivated Arg7 and its wild ancestor W14 were compared. The gene expression counts were normalized to the fragments per kb per million reads (FPKM) for each transcript from the eight libraries (five of Arg7 and three of W14). Transcripts with FPKM < 3 exhibited low abundance in some tissues and were considered silent, in accordance with previous work on white lupin^[19]. Among the 41302 transcripts assembled, 25961 had an FPKM \geq 3 in at least one library and were considered to be transcriptionally active. A total of 2117 and 1963 transcripts were found to be differentially expressed (above twofold change, $P < 0.05$, FPKM \geq 3) in the leaf and stem, respectively, between Arg7 and W14, especially whereas 1875, 3584 and 3967 transcripts showed differential expressions comparing the AR90, AR150, AR210 in Arg7 to WR150 in W14. The distribution of these differentially expressed genes is shown in Figs. 2a–2c. It seems more transcripts were differentially expressed in storage roots between Arg7 to W14. In addition, we identified 1662 and 6577 transcripts that exhibited a species-specific expression pattern in Arg7 and W14, respectively (Fig. 2d). All differently expressed genes in leaf, stem and storage roots between Arg7 and W14 have been assigned to 103, 93 and 119 pathways (AR150 vs WR150), respectively (Appendix D).

3.4 Transcripts/genes involved in light response predominantly expressed in cultivated cassava

When we compared the genes transcribed in functional leaves of Arg7 to W14, a total of 2,117 transcripts were differentially expressed between AL and WL; of these 1,514 were more highly expressed in AL than in WL, and conversely 603 were more highly expressed in WL compared to AL. The significantly differentially expressed genes are shown in Appendix B (Fig. S3). A total of 81 transcripts were much more highly expressed and 23 were expressed at a lower level in AL as compared with WL (Fig. 3).

GO annotation revealed that the most of the transcripts significantly expressed in AL are genes involved in photosynthesis. Fifty five of 81 upregulated genes focus on light reactions, photosystem I (PSA) and photosystem II (PSB). The two light-harvesting complex families, namely,

LHC-I with six gene members had 2.8- to 42.6-fold expression, and LHC-II with another six members had 2.7- to 7.9-fold expression. Ten PSA subunits had 2.5-fold to 49.1-fold and 14 PSB subunits had 3.1- to 45.1-fold higher expression in Arg7. Furthermore, genes in the electron carrier, electron transporter and ATP synthase families were 5.7-, 4.2- and 3.4-fold more highly expressed in AL than WL. 17 of 84 upregulated transcripts in Arg7 belonged to the Calvin cycle of carbon fixation, such as genes for Rubisco small subunit, which has a regulatory function, but not the Rubisco large subunit with a catalytic function, NADPH, 3-PGA, GAP and PEPC. These genes were expressed 3–6 fold higher in Arg7 compared with W14 (Fig. 3). The expression level of genes for ATPase III, Rubisco S and PEPC in AL was validated by quantitative reverse transcription PCR (Fig. 4). A few of the genes in the photorespiration pathway had a higher expression level in Arg7, such as one member of the glycolate oxidase family, glycine decarboxylase, hydroxypyruvate reductase and phosphoribulokinase/uridine kinase were expressed 6.9-, 5.5-, 13.5- and 10.3-fold, respectively, in comparison to W14. Only nine genes had significantly lower expression in Arg7 than in W14, being members of photosystem II, such as PSBA, PSBB, PSBC, PSBD, and PSBZ, and two in photosystem I (PSAA, PSAB) and a subunit of cytochrome B6, that ancestral subunits maybe not yet been selected in process of domestication cassava (Appendix B, Fig. S2). These gene expression patterns, involving light response and carbon fixation, are consistent with cultivated cassava being a C₃–C₄ plant in photosynthesis as reported^[1,2].

3.5 Transcripts/genes for transportation and sucrose degradation with upregulated expression in developing stem of cassava

A total of 1963 transcripts were differentially expressed between Arg7 (AS) and W14 stems (WS); 1195 of these had significantly higher expression in AS than in WS, whereas 768 genes conversely showed lower expression. Eighteen gene families with 83 members involved in transport were significantly upregulated and another 17 families with 75 members were downregulated in stem of Arg7 compared with in W14. It is notable that the ABC transporters and ATPase relative transporters, potassium channel (AKT1, ATKC1, KAT3), transporters for calcium (calmodulin binding) and other unspecified cations, sugars, peptides, amino acids had 8- to 35.6-fold higher expression in developing stems of Arg7 than W14. Some members of transporter gene families, such as water channels and transport of nitrates (NPT2), phosphate, and amino acids had low expression in stems of Arg7 but were significantly higher in W14 (Fig. 5; Appendix E), indicating that the expression of these genes have been reduced through domestication. In the glycolysis pathway, the process of glucose degradation seems to have been strengthened.

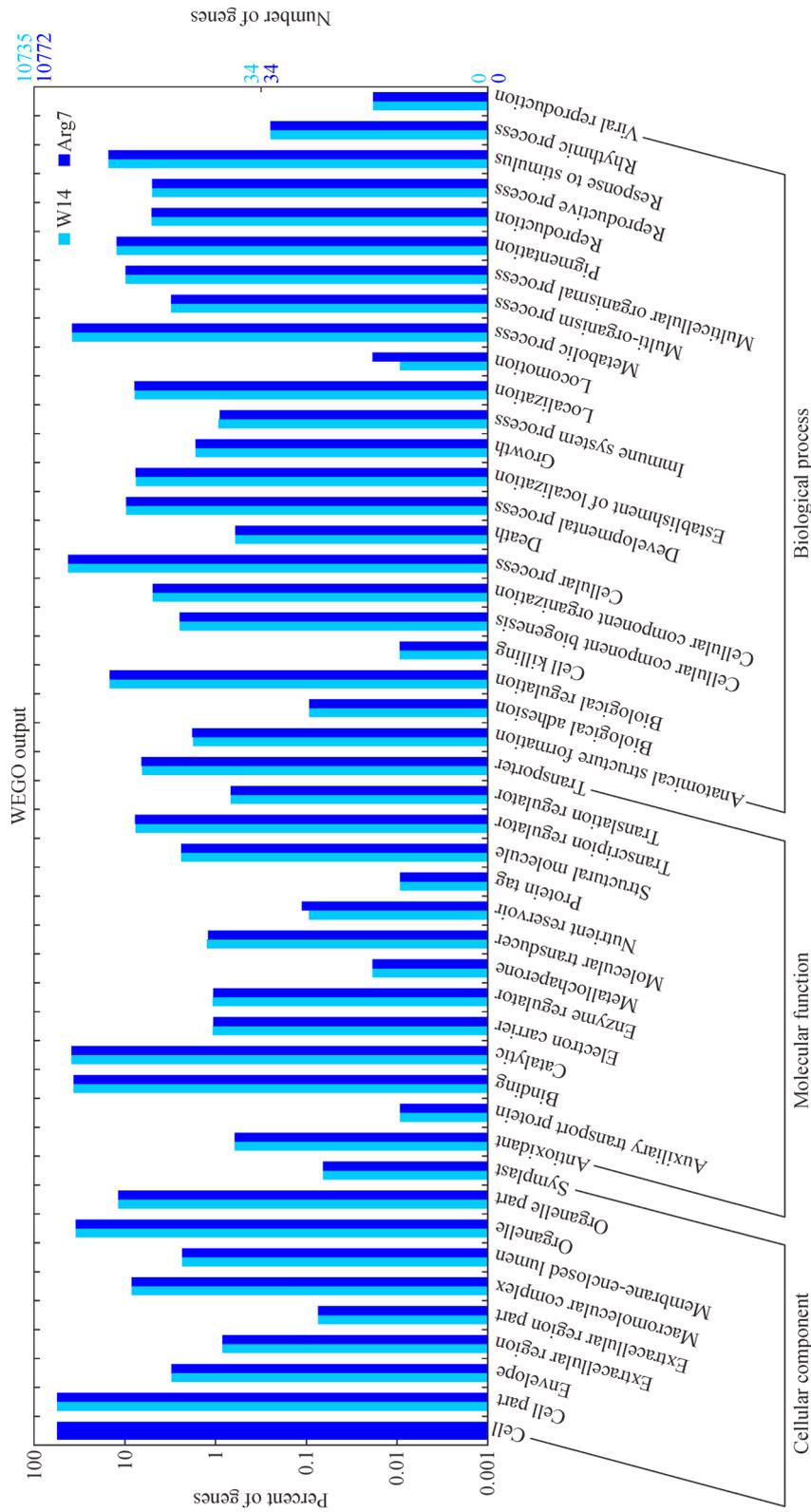


Fig. 1 Gene Ontology (GO) assignments for the transcriptome in leaves and storage roots of Arg7 and W14. Summarized are the enriched genes in three main GO categories: biological process, cellular component, and molecular function. The left y-axis indicates the percentage of genes in each category. The right y-axis indicates the number of genes in the same category.

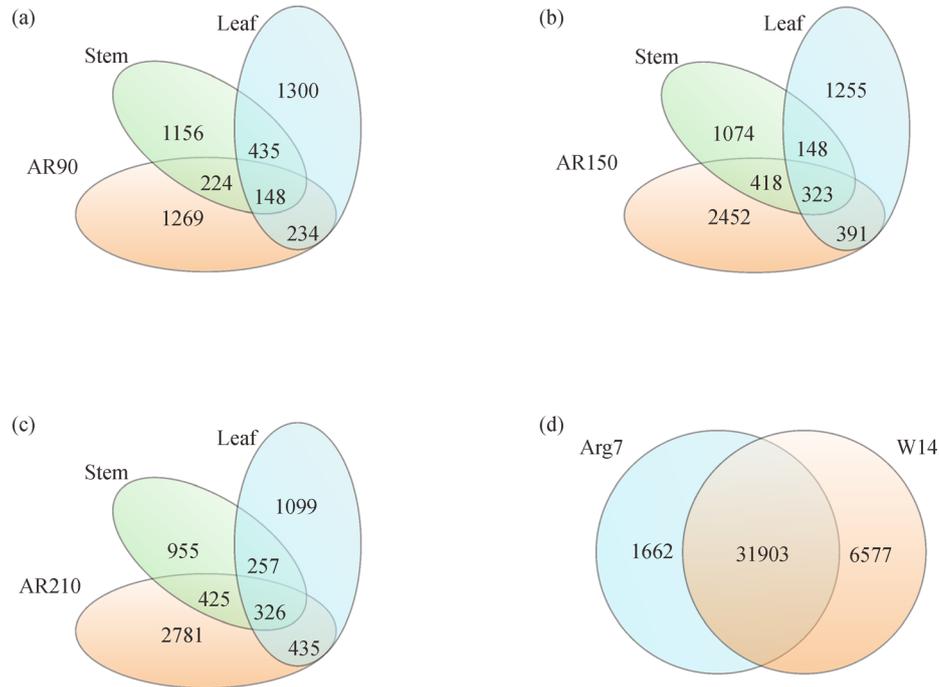


Fig. 2 Number of genes expressed differentially in leaves, stems, and storage roots between Arg7 and W14. (a–c) Distribution of the genes expressed differentially in leaf, stem, and storage root between Arg7 and W14; AR90, AR150, and AR210 refer to Arg7 storage roots at three growth stages compared with WR150; (d) the numbers of species-specific genes and genes shared between Arg7 compared to W14.

Genes for several key rate-limiting enzymes such as PFK that consumes ATP, PGK and PK that both produce ATP, GAPDH producing NADPH reducing power, and ALDO were predominantly expressed in cultivated Arg7 compared with W14 (Fig. 6; Appendix E). Meanwhile, There were 30 two families of transcription factor with 212 members with > 3-fold higher in expression in stems of Arg7 than in W14, such as MYB, HSF, AP2-EREBPs, G2 like and unspecified DNA binding proteins (Appendix B, Fig. S3; Appendix E) that may respond to environmental factors and regulate the marked changes in metabolic processes that has occurred during the domestication of cassava.

3.6 Transcripts/genes involved in key steps of starch synthesis and energy transport were predominantly expressed in developing storage root of cassava

Among the differential transcripts in developing storage root, we focused on the variation of transcripts/genes for starch metabolism and molecular transport pathways, and the transcription factors involved. Among the three comparative transcriptome results for AR90, AR150 and AR210 vs WR150, the gene expression patterns were similar (Appendix A). It was shown that 17 genes in the starch synthesis pathway had 3- to 40-fold higher expression in AR150 than in WR150. They were only involved in three key restricted steps. (1) Sucrose

degradation; the first step in starch synthesis, e.g, genes for six invertases (INVs) and fructose-1,6-bisphosphatase (FBP) with 4- to 11.5-fold greater expression in Arg7 than in W14. (2) Starch synthesis; genes for starch synthesis (GBSS) and glucan phosphorylase (GlgP) had markedly higher expression, 23.8- and 9.1-fold, respectively, in AR150 than WR150. Genes for starch branching enzyme (SBE2.1), glucose-1-phosphate adenylyltransferase (APL3) also had 3- to 3.2-fold higher expression than W14. The high expression of GBSS1 and SBE2.2 in storage root of Arg7 compared to W14 was confirmed by qRT-PCR (Fig. 7). (3) Starch degradation; we found five genes involved in starch hydrolysis had significantly higher expression in AR150 than in WR150, such as the gene for maltose transmembrane transporter (RCP1) with 37.9-fold higher expression than in W14, and the expression of genes for three α -amylases and β -amylase were 6- to 13-fold higher than in W14 (Fig. 8). This indicates that there are robust chemical processes demanding energy and sources materials. For transportation, a total of 109 genes for transporters were significantly upregulated. Predominantly these are involved in energy transport; 13 members of ABC (ATP binding cassette) transporters, which utilize the energy of ATP binding and hydrolysis to energize the translocation of various substrates across membranes, had significantly higher (4- to 32-fold) expression in AR150 than in WR150. In particular, the genes for the H^+ transporting

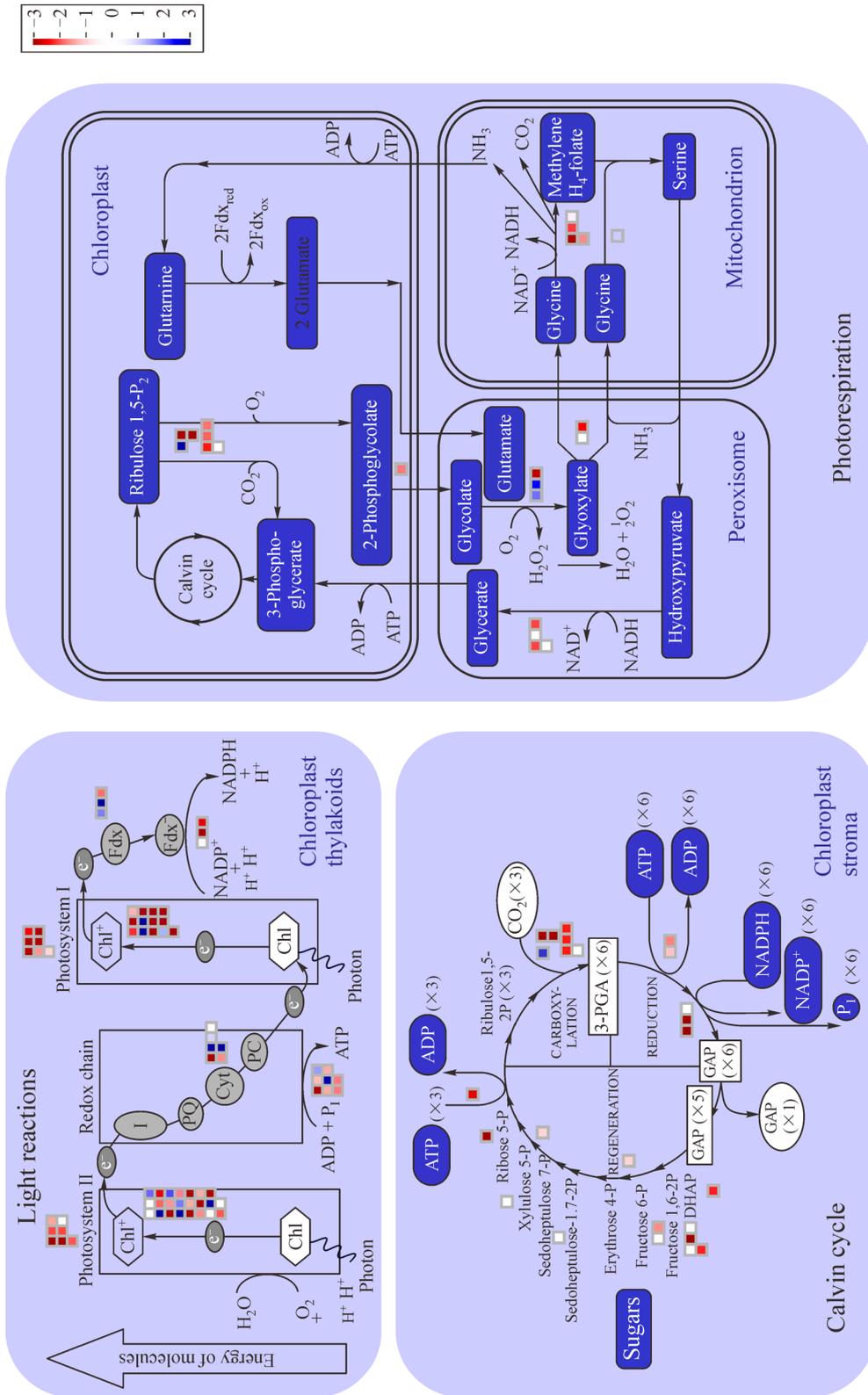


Fig. 3 Genes predominantly expressed in light reactions, Calvin cycle but not in photorespiration pathways in leave of cultivated Arg7 compared to n wild W14. Colored rectangles correspond to the value of log₂ (the FPKM of WL/AL); blue means that a gene expressed folds in Arg7 compared to W14, but red is converse.

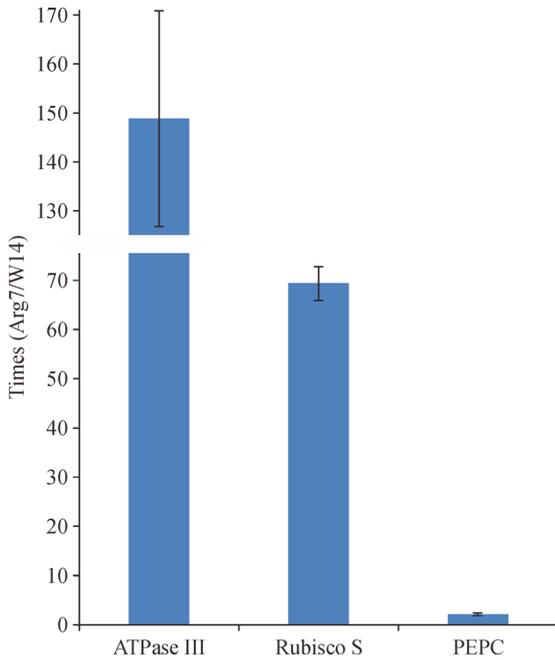


Fig. 4 Validation of expression level of part of genes in leaves of cassava plant, as ATPase III, Rubisco small subunit, and PEPCase by RT-PCR, the expression level 148-, 68- to 3-fold in Arg7 compared to W14

pyrophosphatases (AVP1, AVP3) absolutely high expressed in Arg7 and had a 17-fold higher than in W14, indicating an enhanced requirement for transport, and also serving as a sources of inorganic phosphate used in other biological processes. Six genes for p-ATPase and v-ATPase (AHA4) had 18- to 20.5-fold higher expression in Arg7 and genes for 13 mitochondrial solute carriers (misc transporters), and 14 metabolite transporters at the mitochondrial membrane were expressed 5- to 68-fold higher than in W14. These proteins generally catalyze neutral exchanges, but some are electrogenic or involve H⁺ co-transport, where the energy center has diverse metabolic functions. Among them, NAT7, which belongs to the nucleobase/ascorbate transporter family, was predominant, being 26.8-fold more highly expressed in Arg7 than W14. It is located in the plasma membrane and plasmodesmata, and may be involved in efficient transportation of molecules between cells in storage root of cassava. The function of NAT7 requires more attention. Three genes for water channel proteins (PIP, TIP) also had significantly higher expression. Another 56 upregulated transcripts/genes, including transporters for nitrates (NRT, NTP3), phosphate (PHT4-5), potassium (KUP7), sulfate (SULTR2, SULTR5), amino acids, nucleotides and peptides, and sugars (TMT1, ZIF1). These also include the

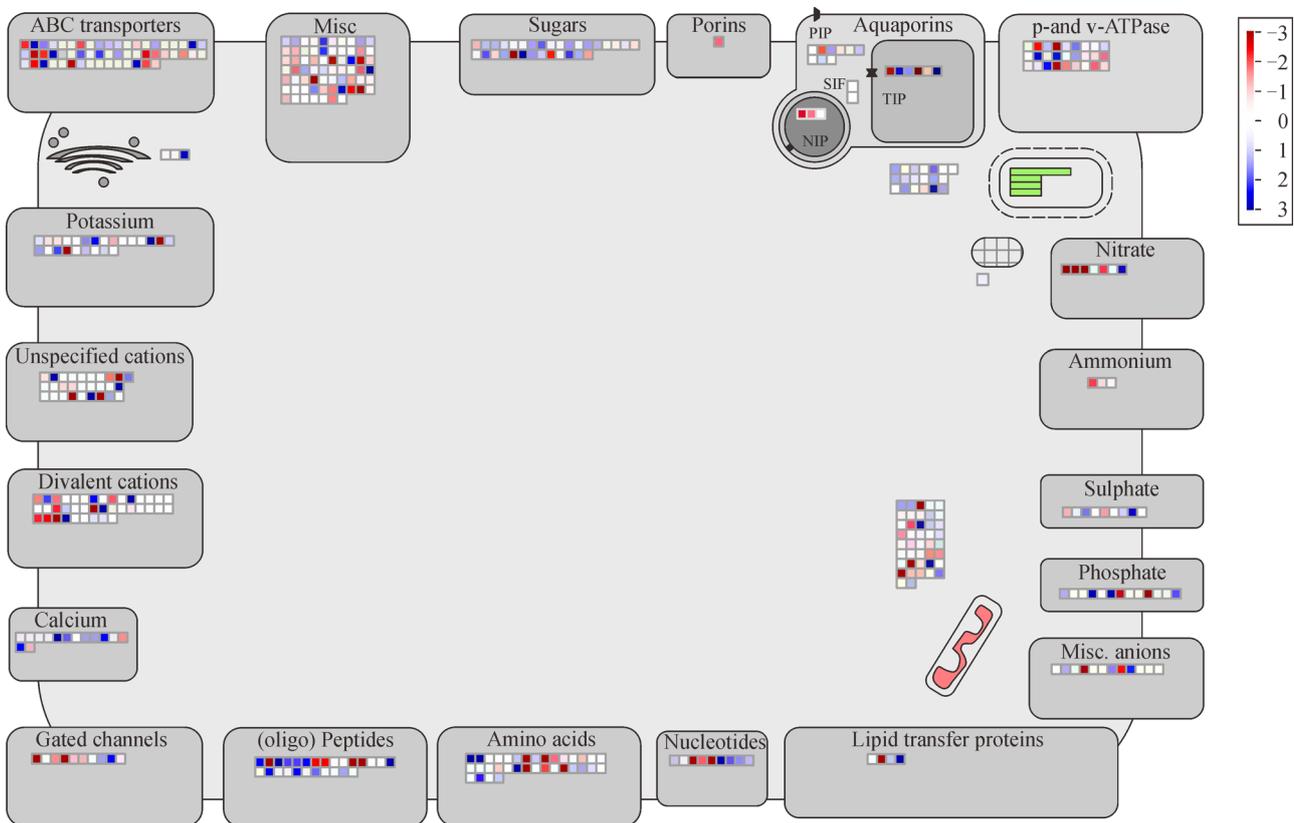


Fig. 5 Genes involved in transport predominantly expressed in developing stems of cultivated Arg7 compared to W14. Colored rectangles correspond to the value of log₂ (the FPKM of WS/AS): blue means that a gene expressed folds in W14 compared to Arg7, but red is converse.

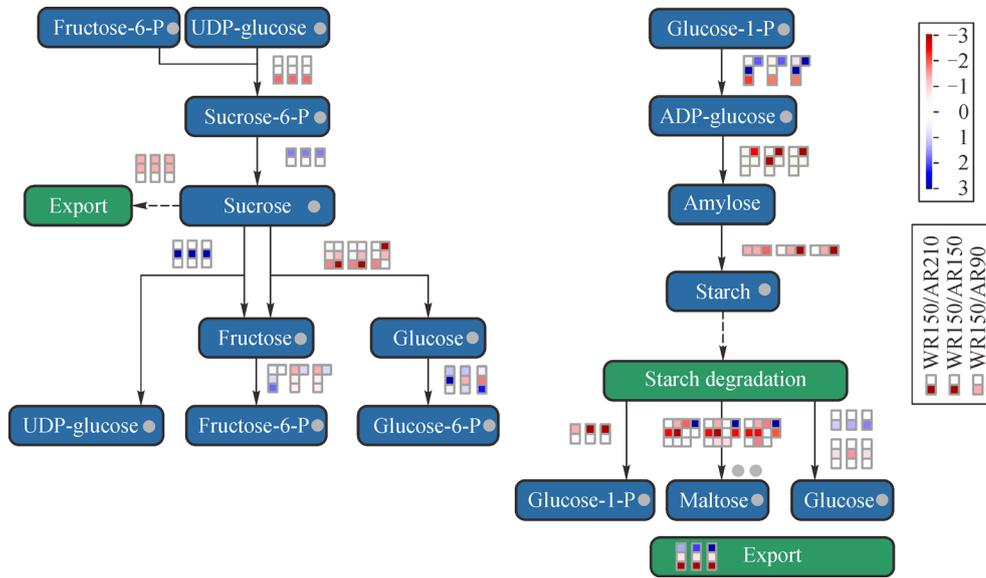


Fig. 8 Differentially expressed genes in the pathways of sucrose and starch metabolism with three development stage of storage root. Colored rectangles correspond to the value of \log_2 (the FPKM of a gene in W14/that in Arg7): blue means that a gene expressed folds in W14 compared to Arg7, whereas red is converse

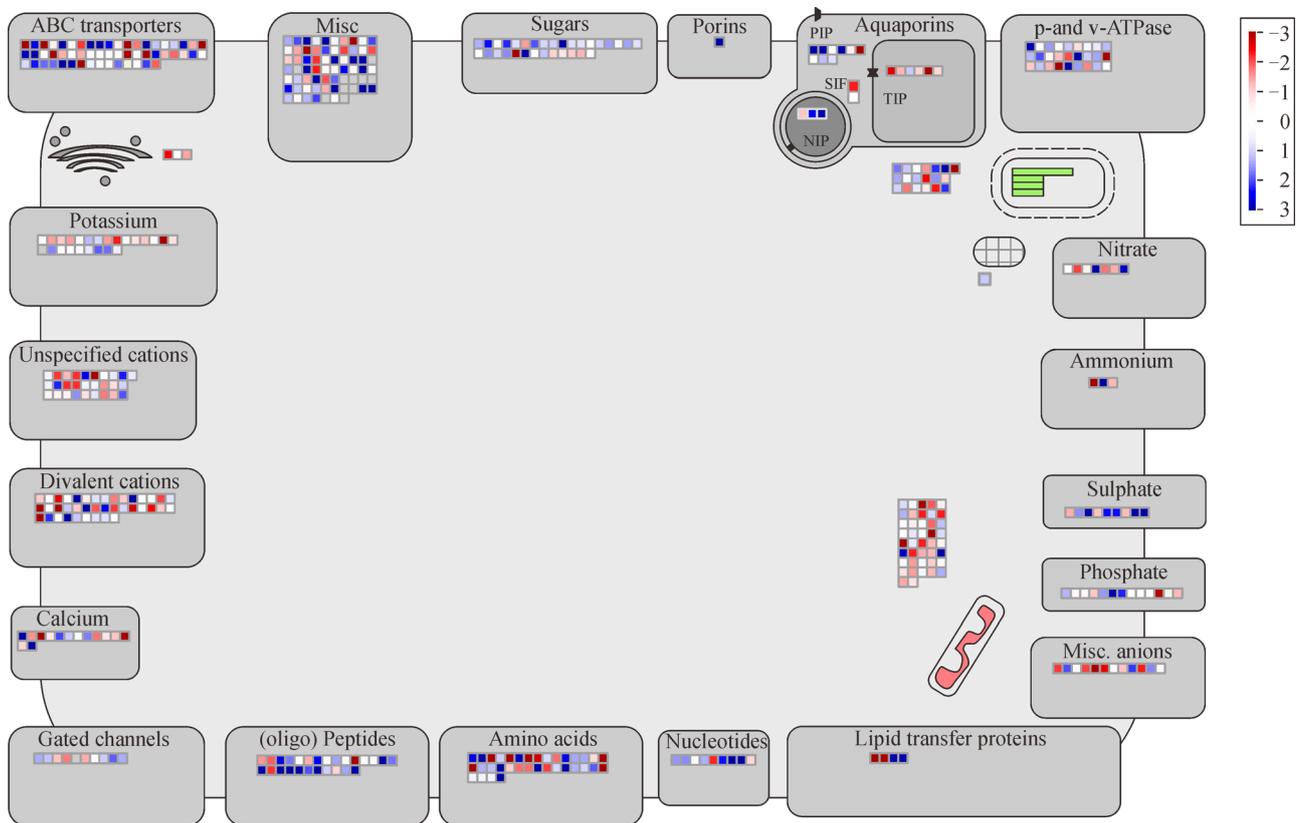


Fig. 9 Genes involved in transport predominantly expressed in storage roots of cultivated Arg7 compared to W14. Colored rectangles correspond to the value of \log_2 (the FPKM of WR150/AR150): blue means that a gene expressed folds in W14 compared to Arg7, but red is converse.

sub-networks with hub genes of starch synthase (GBSS), glucan phosphorylase (GlgP), GBS2.1 and GBS2.2. Seven genes for energy transporters, including ABC, ALDO, pantothenate kinase (PANK2), CXIP2 and UBX domain-containing protein, and 15 genes for transcription factors, including HCF, MYB, AGL44, ERF9, HB54, NAC2, SCL1 and protein binding, have been integrated into the network (Appendix B, Fig. S6). This verified their expression as being positively correlated with the expression of key genes in starch biosynthesis.

4 Discussion

4.1 Enhancing expression of photosynthetic light reactions genes was a vital event in domestication of cassava

It was shown that although there are 1514 genes highly expressed in leaves of the cassava cultivars, the remarkable finding is the genes for light reactions such as photosystem I (PSA) and photosystem II (PSB), the two light-harvesting complex families and other members in the light reaction chains, had significantly higher expression, along with some key genes in Calvin cycle of carbon fixation (RUBP, NAPDH, 3-PGA, GAP and PEPC). Cultivated cassava normally has a high net photosynthesis rate, up to $54 \text{ mg} \cdot \text{dm}^{-2} \cdot \text{h}$ under certain conditions^[31]. It is considered as a C_3 – C_4 intermediate plant, whereas the wild progenitor (*M. esculenta* ssp. *flabellifolia*) exhibits C_3 characteristics^[1,2]. It has been reported that photorespiration-related genes had higher expression in the C_3 – C_4 intermediate species *Flaveria ramosissima*^[32] but not in cassava. There have been reports of increased response to light intensity under drought stress^[33] and shade mainly induced expression of genes for light reactions, light-harvesting complexes as well as light signaling in cassava leaf^[34]. Comparative genome annotation of W14 and KU50 has identified genes enriched for response to stimuli including light, temperature, water and oxygen evolution^[24], but our research further illuminated the specific gene and gene families that respond to light which have been selectively enhanced during domestication. These numerous characteristics in light response and carbon fixation support the conclusion that cultivated cassava is a C_3 – C_4 plant.

4.2 Expression of active sucrose degradation and transport genes indicates robust energy metabolism in stem of cassava

There was significantly higher expression of genes for several key rate-limiting enzymes in the glycolysis pathway (PFK, PGK and PK, GAPDH and ALDO) in stems of domesticated cassava compared with W14, as well as 212 members transcription factor families. According to information from biochemical and transcriptome analysis of stem phloem sap, there is a powerful stem transport

capacity which may affect starch accumulation in the storage roots^[35]. Cassava stems are used for reproduction and to keep the plant upright, whereas the roots store materials, which differs from most plants where bulbs or tubers are used for material storage and reproduction^[21]. This unique feature of cassava is not well understood.

4.3 Expression of rate-limiting genes in starch synthesis and energy transport are significantly enhanced in storage root of cassava

Based on comparison of three developmental stages of the storage roots of Arg7 with wild W14, we found that the genes specially involved in three key restricted steps, sucrose degradation (INVs, FBP), starch synthesis (GBSS, GlgP, SBE2, APL3) and starch degradation (RCP1, α -AMY, β -AMY) had significantly higher expression in the cultivar than the wild ancestor. Also, robust energy transport activity in developing storage root was observed, predominantly in selected ABC (ATP binding cassette) transporters, H^+ transporting pyrophosphatase (AVP1, AVP3), p-ATPase and v-ATPase (AHA4), mitochondrial solute carriers (misc transporters) and many metabolite transporters at the mitochondrial membrane. Arg7 is a high starch content cultivar with 30% starch in mature fresh storage roots, but W14 has a low starch content at less than 5%. At the genome level, it has been reported that expansion of copy number and alternative splicing were found in genes for some key steps in starch synthesis, such as sucrose transporter (SUTs), sucrose synthases (GBSS) and ADP glucose pyrophosphorylase (APL)^[24]. It was recently reported from a proteomics study that the enzymes involved in energy metabolism and binding activity, such as glucan phosphorylase, ADP-glucose pyrophosphorylase, ATPase, fructokinase, malate dehydrogenase and multiple enzymes participated in binding, may have important roles in cassava storage roots^[36]. The above evidence and other reports in wheat and other crops^[22,37,38], strongly support our findings.

5 Conclusions

Large-scale comparative transcriptome analysis revealed the genes important in leaf, stem and storage root of cultivated cassava by comparison with its wild ancestor. This new understanding focused attention on specific genes for light harvesting and response, together with Calvin cycle genes in leaf, other genes for transport and key rate-limiting enzymes coupling ATP consumption in the glycolysis pathway in stem, and genes for sucrose degradation, amylose synthesis and hydrolysis. The expression of genes involved in three key steps in starch synthesis associated with storage has been enhanced during domestication. The co-expression network in each organ linked important transcription factors to their

putative targets, and further confirmed some of important hub genes in photosynthesis, transportation and starch biosynthesis. This advance will contribute to elucidating the mechanism of highly efficient carbohydrates accumulation and also genetic improvement in cassava.

Acknowledgements This work was supported by the National Natural Science Foundation of China (31261140363, 31171230), the National Basic Research and Development Program (2010CB126601), China Agriculture Research System (CARS-12wwq), and the Hainan Province Innovative Research Team Foundation (2016CXTD013).

Supplementary materials The RNA-seq reads are deposited in the GenBank/SRR sequence read archive under the accession codes: SRR1298999, SRR1298998, SRR1298996, SRR1299008, SRR1299009, SRR1299005, SRR1299006 and SRR1299007. The online version of this article at <http://dx.doi.org/10.15302/J-FASE-2016045> contains supplementary materials (Appendixes A–E).

Compliance with ethics guidelines Zhiqiang Xia, Xin Chen, Cheng Lu, Meiling Zou, Shujuan Wang, Yang Zhang, Kun Pan, Xincheng Zhou, Haiyan Wang, and Wenquan Wang declare that they have no conflict of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Angelov M N, Sun J, Byrd G T, Brown R H, Black C C. Novel characteristics of cassava, *Manihot esculenta* Crantz, a reputed C₃–C₄ intermediate photosynthesis species. *Photosynthesis Research*, 1993, **38**(1): 61–72
2. Calatayud P A, Barón C H, Velásquez H, Arroyave J A, Lamaze T. Wild *Manihot* species do not possess C₄ photosynthesis. *Annals of Botany*, 2002, **89**(1): 125–127
3. Olsen K M, Schaal B A. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**(10): 5586–5591
4. Olsen K M, Schaal B A. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. *American Journal of Botany*, 2001, **88**(1): 131–142
5. Colombo C, Second G, Charrier A. Genetic relatedness between cassava (*Manihot esculenta* Crantz) and *M. flabellifolia* and *M. Peruviana* based on both RAPD and AFLP markers. *Genetics and Molecular Biology*, 2000, **23**(2): 417–423
6. Olsen K M. SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology*, 2004, **56**(4): 517–526
7. Pujol B, Mühlén G, Garwood N, Horoszowski Y, Douzery E J, McKey D. Evolution under domestication: contrasting functional morphology of seedlings in domesticated cassava and its closest wild relatives. *New Phytologist*, 2005, **166**(1): 305–318
8. Prochnik S, Marri P R, Desany B, Rabinowicz P D, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar D S, Rounsley S. The cassava genome: current progress, future directions. *Tropical Plant Biology*, 2012, **5**(1): 88–94
9. Li Y Z, Pan Y H, Sun C B, Dong H T, Luo X L, Wang Z Q, Tang J L, Chen B. An ordered EST catalogue and gene expression profiles of cassava (*Manihot esculenta*) at key growth stages. *Plant Molecular Biology*, 2010, **74**(6): 573–590
10. Verdier V, Restrepo S, Mosquera G, Jorge V, Lopez C. Recent progress in the characterization of molecular determinants in the *Xanthomonas axonopodis* pv. *manihotis*-cassava interaction. *Plant Molecular Biology*, 2004, **56**(4): 573–584
11. Kemp B P, Beeching J R, Cooper R M. cDNA-AFLP reveals genes differentially expressed during the hypersensitive response of cassava. *Molecular Plant Pathology*, 2005, **6**(2): 113–123
12. Sojikul P, Kongsawadworakul P, Viboonjun U, Thaiprasit J, Intawong B, Narangajavana J, Svasti J. AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. *Physiologia Plantarum*, 2010, **140**(2): 189–298
13. Lopez C, Soto M, Restrepo S, Piégu B, Cooke R, Delseny M, Tohme J, Verdier V. Gene expression profile in response to *Xanthomonas axonopodis* pv. *manihotis* infection in cassava using a cDNA microarray. *Plant Molecular Biology*, 2005, **57**(3): 393–410
14. Utsumi Y, Tanaka M, Morosawa T, Kurotani A, Yoshida T, Mochida K, Matsui A, Umemura Y, Ishitani M, Shinozaki K, Sakurai T, Seki M. Transcriptome analysis using a high-density oligomicroarray under drought stress in various genotypes of cassava: an important tropical crop. *DNA Research*, 2012, **19**(4): 335–345
15. An D, Yang J, Zhang P. Transcriptome profiling of low temperature-treated cassava apical shoots showed dynamic responses of tropical plant to cold stress. *BMC Genomics*, 2012, **13**(1): 64
16. Reilly K, Bernal D, Cortes D F, Gomez-Vasquez R, Tohme J, Beeching J R. Towards identifying the full set of genes expressed during cassava post-harvest physiological deterioration. *Plant Molecular Biology*, 2007, **64**(1–2): 187–203
17. Gao Y, Xu H, Shen Y, Wang J. Transcriptomic analysis of rice (*Oryza sativa*) endosperm using the RNA-Seq technique. *Plant Molecular Biology*, 2013, **81**(4–5): 363–378
18. Oono Y, Kobayashi F, Kawahara Y, Yazawa T, Handa H, Itoh T, Matsumoto T. Characterization of the wheat (*Triticum aestivum* L.) transcriptome by *de novo* assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC Genomics*, 2013, **14**(1): 77
19. O'Rourke J A, Yang S S, Miller S S, Bucciarelli B, Liu J, Rydeen A, Bozsoki Z, Uhde-Stone C, Tu Z J, Allan D, Gronwald J W, Vance C P. An RNA-Seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. *Plant Physiology*, 2013, **161**(2): 705–724
20. Allie F, Pierce E J, Okoniewski M J, Rey C. Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genomics*, 2014, **15**(1): 1006
21. El-Sharkawy M A. Cassava biology and physiology. *Plant Molecular Biology*, 2004, **56**(4): 481–501
22. McMaugh S J, Thistleton J L, Anschaw E, Luo J, Konik-Rose C, Wang H, Huang M, Larroque O, Regina A, Jobling S A, Morell M

- K, Li Z. Suppression of starch synthase I expression affects the granule morphology and granule size and fine structure of starch in wheat endosperm. *Journal of Experimental Botany*, 2014, **65**(8): 2189–2201
23. Tetlow I J, Morell M K, Emes M J. Recent developments in understanding the regulation of starch metabolism in higher plants. *Journal of Experimental Botany*, 2004, **55**(406): 2131–2145
24. Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Meller B L, Zhang P, Luo M C, Xiao G, Liu J, Yang J, Chen S, Rabinowicz P D, Chen X, Zhang H B, Ceballos H, Lou Q, Zou M, Carvalho L J, Zeng C, Xia J, Sun S, Fu Y, Wang H, Lu C, Ruan M, Zhou S, Wu Z, Liu H, Kannangara R M, Jorgensen K, Neale R L, Bonde M, Heinz N, Zhu W, Wang S, Zhang Y, Pan K, Wen M, Ma P A, Li Z, Hu M, Liao W, Hu W, Zhang S, Pei J, Guo A, Guo J, Zhang J, Zhang Z, Ye J, Ou W, Ma Y, Liu X, Tallon L J, Galens K, Ott S, Huang J, Xue J, An F, Yao Q, Lu X, Fregene M, Lopez-Lavalle L A, Wu J, You F M, Chen M, Hu S, Wu G, Zhong S, Ling P, Chen Y, Wang Q, Liu G, Liu B, Li K, Peng M. Cassava genome from a wild ancestor to cultivated varieties. *Nature Communications*, 2014, **10** (5): 5110
25. Zhong S L, Joung J G, Zheng Y, Chen Y R, Liu B, Shao Y, Xiang J Z, Fei Z J, Giovannoni J J. High-throughput Illumina strand-specific RNA sequencing library preparation. *Cold Spring Harbor Protocols*, 2011, **2011**(8): 940–949
26. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley D R, Pimentel H, Salzberg S L, Rinn J L, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 2012, **7**(3): 562–578
27. Quevillon, E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Research*, 2005, **33**(Suppl. 2): W116–W120
28. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, 2006, **34**(Suppl. 2): W293–W297
29. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 2008, **36**(Suppl. 1): D480–D484
30. Patrick E, Meyer, Frederic L, Gianluca B. minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, 2008, **9**(1): 461
31. Engelmann S, Bläsing O E, Gowik U, Svensson P, Westhoff P. Molecular evolution of C₄ phosphoenol-pyruvate carboxylase in the genus *Flaveria*—a gradual increase from C₃ to C₄ characteristics. *Planta*, 2003, **217**(5): 717–725
32. Gowik U, Bräutigam A, Weber K L, Weber A P, Westhoff P. Evolution of C₄ photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C₄? *Plant Cell*, 2011, **23**(6): 2087–2105
33. Sakurai T, Plata G, Rodriguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M. Sequencing analysis of 20000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology*, 2007, **7**(1): 66
34. Ding Z H, Zhang Y, Xiao Y, Liu F F, Wang M H, Zhu X G, Liu P, Sun Q, Wang W Q, Peng M, Tom Brutnell & Li P H. Transcriptome response of cassava leaves under natural shade. *Scientific Reports*, 2016, **6**: 31673
35. Li Y Z, Zhao J Y, Wu S M, Fan X W, Luo X L, Chen B S. Characters related to higher starch accumulation in cassava storage roots. *Scientific Reports*, 2016, **6**: 19823
36. Wang X C, Chang L L, Tong Z, Wang D Y, Yin Q, Wang D, Jin X, Yang Q, Wang L M, Sun Y, Huang Q X, Guo A P, Peng M. Proteomics profiling reveals carbohydrate metabolic enzymes and 14–3–3 proteins play important roles for starch accumulation during cassava root tuberization. *Scientific Reports*, 2016, **6**: 19643
37. Kang G, Liu G, Peng X, Wei L, Wang C, Zhu Y, Ma Y, Jiang Y, Guo T. Increasing the starch content and grain weight of common wheat by over-expression of the cytosolic AGPase large subunit gene. *Plant Physiology & Biochemistry Ppb*, 2013, **73**(6): 93–98
38. Asare E K, Båga M, Rossnagel B G, Chibbar R N. Polymorphism in the barley granule-bound starch synthase 1 (*gbss1*) gene associated with grain starch variant amylose concentration. *Journal of Agricultural and Food Chemistry*, 2012, **60**(40): 10082–10092